BILAN D'UN LABORATOIRE D'UTILISABILITÉ DES INTERFACES HUMAIN-MACHINE[1]

Jean-Marc ROBERT (1), Daniel ENGELBERG (2)

(1) Ecole Polytechnique de Montréal, Département de mathématiques et de génie industriel, Case postale 6079, Succ. centre-ville, Montréal, Québec H3C 3A7
(2) CRIM, 1801 avenue McGill College, Bureau 800, Montréal, Québec H3A 2N4

RESUME

Les laboratoires d'utilisabilité ont pour rôle d'évaluer la qualité des interfaces humain-machine en faisant appel à des utilisateurs à qui on demande d'effectuer des tâches pendant lesquelles on note les problèmes de l'interface et on mesure la performance et la satisfaction humaine. Cet article trace le bilan du laboratoire d'utilisabilité ICONE qui est opérationnel depuis un an dans différents milieux industriels. Il présente la méthodologie suivie dans ce laboratoire, soient les objectifs, les tâches, les utilisateurs, l'interface et la procédure de tests. Les principaux résultats sont les suivants. Les tests sont tous effectués sur le terrain plutôt qu'en laboratoire. Plusieurs sessions de tests sont nécessaires pour chaque projet; au tout début, elles sont axées sur les problèmes de l'interface, par la suite elles visent à mesurer la performance et la satisfaction des utilisateurs. Les tests portent surtout sur la facilité d'utilisation de l'interface à court terme. Ils s'avèrent un excellent moyen de compléter l'analyse des besoins et d'analyser l'activité des utilisateurs aux prises avec l'interface.

CONTEXTE ET OBJECTIFS

Depuis le milieu des années 80, on assiste à la création de nombreux laboratoires d'utilisabilité des interfaces humain-machine à travers le monde (voir le Centre de compétence en ergonomie de Bayonne 1993; Chignell 1992; ICONE 1994; Nielsen 1994a; LabiUtil 1995; Reed, 1992). Il convient alors de définir les notions d'utilisabilité et de laboratoire d'utilisabilité, puis de faire le bilan de l'un de ces laboratoires, ICONE, qui est opérationnel depuis un an dans différents milieux.

Utilisabilité. La notion d'utilisabilité (traduction de *usability*) est multidimensionnelle. Elle signifie qu'un produit ou qu'un système est facile et agréable à apprendre et à utiliser. A la base, deux aspects sont donc en cause: la performance et la satisfaction des utilisateurs. La performance peut être liée à l'apprentissage ou à l'utilisation du système. On mesure la facilité d'apprentissage à travers le temps d'apprentissage et le taux de rétention de l'utilisateur, et la facilité d'utilisation à travers la productivité et le taux d'erreurs des utilisateurs. La facilité d'apprentissage et la facilité d'utilisation peuvent être en opposition car un système peut être facile à apprendre tout en étant mauvais dans son utilisation. Il faut donc choisir l'objectif à privilégier ou faire le juste compromis entre les deux.

Laboratoire d'utilisabilité. Un laboratoire d'utilisabilité est un endroit où l'on teste la qualité des interfaces du point de vue humain en demandant à de futurs utilisateurs d'effectuer un ensemble de tâches pré-définies pendant lesquelles on identifie les problèmes de l'interface et on mesure la performance et la satisfaction humaine. L'interface est donc évaluée par rapport à certains utilisateurs et certaines tâches, de sorte que la portée de l'évaluation est clairement établie. Le but des tests d'utilisabilité est d'améliorer la qualité des interfaces humain-machine.

---

[1] In Compte-rendu du Congrès international de génie industriel de Montréal, 18-20 octobre 1995, Montréal, Québec.

Avantages. Les interfaces de meilleure qualité qui découlent des tests d'utilisabilité représentent de nombreux avantages pour les utilisateurs, pour les entreprises qui achètent des logiciels et celles qui les développent. Pour les utilisateurs, les avantages consistent en une réduction du temps d'apprentissage, du temps d'exécution des tâches et du nombre d'erreurs, une plus grande autonomie avec le système et une plus grande satisfaction. Pour les entreprises qui achètent des logiciels, les avantages se traduisent par une diminution du temps de formation des employés et du nombre d'erreurs, une plus grande qualité du travail accompli et une meilleure productivité. Enfin, pour les entreprises qui développent et vendent des logiciels, les avantages se traduisent par de meilleures revues de presse et une bonne image de marque (qui devraient favoriser les ventes), et une diminution des demandes de services après vente, des exigences de formation des clients et du volume de la documentation accompagnant l'interface.

Place des laboratoires. Le recours au laboratoire d'utilisabilité s'inscrit dans le cadre d'une approche de conception centrée sur l'utilisateur et sa tâche (ex., Gould, 1988; Mayhew, 1992; Robert & Fiset, 1992). Il fait partie d'une démarche itérative de conception de systèmes qui s'appuie fortement sur le prototypage et qui consiste à tester l'interface auprès des utilisateurs de façon continue et très tôt dans le processus de conception, à la modifier et la retester jusqu'à ce qu'elle soit jugée satisfaisante. Les résultats sont très probants: Whiteside et al. (1988) estiment que l'on peut améliorer l'utilisabilité de 30% à chaque itération. On procède ainsi parce que l'on ne dispose pas à l'heure actuelle de modèles théoriques assez puissants pour prédire la performance humaine face à une interface ou encore parce qu'on est incapable de façon intuitive de détecter les failles d'une interface ou d'anticiper la performance des utilisateurs. L'étude de Tullis (1993) illustre bien cette réalité: elle a montré que 28 développeurs d'interfaces ayant plusieurs années d'expérience étaient incapables, à partir de leur seul jugement, d'ordonner 7 différentes interfaces selon le critère de rapidité d'exécution d'une tâche par les utilisateurs.

Une méthode d'évaluation parmi d'autres. Les tests d'utilisabilité, qui impliquent des sujets, correspondent à une méthode d'évaluation des interfaces parmi d'autres (ex., voir Karat, 1988; Mack & Nielsen, 1993). Les autres méthodes ne font pas appel aux utilisateurs. Elles peuvent être basées sur une liste de points à vérifier (ex., la méthode heuristique) ou encore consister à analyser soigneusement le comportement d'une interface lorsqu'on exécute des tâches avec celle-ci (ex., walkthrough). Ces différentes méthodes ne sont pas mutuellement exclusives.

METHODOLOGIE DES TESTS

La méthodologie des tests d'utilisabilité emprunte beaucoup à celle de la recherche en psychologie expérimentale. Cinq principaux points doivent être définis, soient les objectifs, les tâches, les utilisateurs, l'interface et la procédure des tests incluant les mesures à prendre. Les paragraphes qui suivent décrivent chacun de ces points.

Les objectifs

Il faut d'abord bien définir les objectifs des tests d'utilisabilité car ils influent fortement sur les autres points de la méthodologie, notamment sur les mesures à prendre au cours des tests. Les objectifs changent et se raffinent au cours de la conception, allant du global au tout début vers le particulier à la fin, et du qualitatif vers le quantitatif. Ils sont définis en premier lieu à partir des objectifs du client (par ex., "le système doit pouvoir être appris en moins d'une demi-journée", "on doit pouvoir utiliser le système dès les 5 premières minutes"), puis en tenant compte de la phase de conception, des résultats des tests antérieurs, du temps et du budget disponibles.

Compte tenu de ces points, deux questions de base nous aident à définir les objectifs des tests: Cherche-t-on à identifier les problèmes de l'interface ou à mesurer la performance et la satisfaction des utilisateurs? S'agit-il de la performance liée à l'apprentissage ou à l'utilisation du

système? En conséquence, l'objectif d'une session de tests d'utilisabilité sera qualitatif ou quantitatif, tantôt axé sur les problèmes de l'interface, tantôt sur la mesure de la performance et de la satisfaction des utilisateurs.

Conseils pratiques pour définir les objectifs:
• se donner des objectifs précis et opérationnels;
• éviter de tout tester à la fois;
• faire approuver les objectifs par l'équipe de conception;
• définir les exigences à satisfaire en incluant le temps et le budget requis.

Les tâches

Les tests d'utilisabilité se font au moyen de tâches réelles tirées directement du terrain ou de scénarios de tâches élaborés par le responsable des tests, ou les deux à la fois. Dans tous les cas, on doit pouvoir s'appuyer sur une analyse ergonomique du travail.

Un scénario de tâche correspond à un ensemble cohérent de tâches axé sur un but et typique de ce que l'utilisateur sera appelé à faire avec le logiciel. Il a l'avantage de pouvoir être plus général qu'une tâche spécifique tirée du terrain parce qu'il peut combiner les caractéristiques de plusieurs tâches à la fois. Mais surtout, il permet d'étudier des tâches rares ou critiques qui n'ont pas pu être observées au cours de l'analyse du travail, faute de temps.

Pour les applications informatiques très répandues (ex., traitement de texte, graphisme, chiffrier), on peut utiliser des tâches standard *(benchmark task)* qui permettent non seulement de mesurer la performance de l'utilisateur mais de la comparer avec celle obtenue dans des applications semblables. Voici des exemples pour un système de traitement de texte: intervertir deux paragraphes, mettre le texte en deux colonnes, insérer un tableau. Pour les applications informatiques moins répandues, parfois limitées à un seul milieu de travail ou à une seule entreprise, il faut chaque fois avoir des tâches ou des scénarios de tâches sur mesure.

Quatre principaux critères centrés sur la tâche et non indépendants les uns des autres nous guident dans le choix des tâches ou des scénarios de tâches, soit:
• la nature de la tâche (cela dépend de l'application)
• la fréquence (fréquente vs peu fréquente)
• le degré de criticalité (critique vs non critique)
• le degré de difficulté (facile vs difficile).

Un critère centré sur l'interface nous guide également dans ce choix, soit:
• le niveau de développement de l'interface (qui correspond à ce qui est opérationnel dans l'interface).

D'autres critères centrés sur la tâche, sur l'interface ou sur les deux à la fois peuvent s'ajouter au fur et à mesure de l'évolution des tests. Par ex.:
• la partie de l'interface qui crée des difficultés aux utilisateurs.

Deux approches sont possibles pour définir les tâches ou les scénarios de tâches: soumettre une tâche globale à l'utilisateur pour toute la durée de la session (ex., entre 30 et 60 minutes), ou soumettre plusieurs tâches spécifiques de plus courte durée. On peut privilégier la première approche au tout début des tests, surtout si on connaît encore mal la tâche, puis la deuxième approche lorsque l'on cherche à observer des points précis dans l'interface.

Les tâches ou les scénarios de tâches sont définis par le responsable des tests ou par un expert de la tâche qui collabore étroitement avec lui. On fait toujours appel à un expert lorsque la tâche est complexe ou critique, ou qu'on n'a pas eu le temps de l'analyser en profondeur. Il est toujours recommandé de faire valider les choix de tâches par un expert de la tâche.

Deux critiques, nous suggérant des sujets de recherche, peuvent être formulées à l'endroit des scénarios de tâches: 1) on ne dispose pas de méthode permettant de mesurer précisément leur représentativité, c.-à-d. quelle proportion de la tâche ils couvrent et jusqu'à quel point ils la couvrent bien; 2) il y a un problème de fidélité car, face à une même tâche, deux responsables de tests vont vraisemblablement définir des scénarios différents.

Conseils pratiques pour définir les tâches:
- définir les tâches en fonction des objectifs de l'utilisateur et non pas des caractéristiques de l'interface;
- aller du général vers le particulier d'une session de tests à l'autre;
- faire valider les choix de tâches par un expert de la tâche.

Les utilisateurs

Un critère global doit être respecté pour choisir les sujets devant participer aux tests d'utilisabilité, soit leur représentativité par rapport à la population des futurs utilisateurs de l'interface. Cependant, pour assurer cette représentativité, plusieurs paramètres doivent être considérés, tels que l'âge, le sexe, la langue et la culture, l'occupation, la compétence dans le domaine de la tâche, la compétence technique ou informatique, les habiletés (ex., dactylographier, utiliser une souris), les déficiences sensorielles (ex., daltonisme), les attitudes, etc. (voir Scerbo 1995 pour une grille de description des caractéristiques des utilisateurs). Si les utilisateurs proviennent de l'entreprise même où on développe la nouvelle interface (clientèle interne), on pourra aussi considérer la localisation géographique et l'unité administrative à laquelle ils appartiennent à cause des différences possibles dans la culture et l'expérience technique.

Le nombre d'utilisateurs requis pour participer à chaque session de tests d'utilisabilité est une question délicate à traiter. D'un point de vue statistique, il est avantageux d'avoir le plus grand nombre de sujets possible pour avoir une plus grande confiance dans les résultats. D'un point de vue pratique, des contraintes de temps et de budget nous obligent souvent à rechercher le plus petit nombre de sujets possible. Il y a un compromis à faire. L'étude de Virzi (1992) peut nous aider à prendre une décision. Basée sur l'évaluation de trois différents produits, elle montre que la majorité des problèmes de l'interface sont détectés par les tout premiers sujets (N = 3), que les problèmes les plus importants sont susceptibles d'être détectés avec un faible nombre de sujets (N = 2 ou 3), et que l'augmentation du nombre de sujets accroît la probabilité de détecter les problèmes les moins importants. 5 sujets permettent de détecter presque 100% des problèmes très importants, environ 95% des problèmes moyennement importants et près de 60% des problèmes moins importants. Cet auteur conclut que dans la plupart des situations, 5 sujets sont capables au cours d'une session de détecter environ 80% des problèmes d'utilisabilité les plus critiques.

Conseils pratiques pour choisir les utilisateurs:
- prendre pour acquis que l'utilisateur moyen n'existe pas: par conséquent, avoir des représentants de chaque groupe d'utilisateurs;
- ne pas se contenter des utilisateurs choisis par les représentants de l'entreprise car on est susceptible de se retrouver avec les meilleurs employés seulement, ou ceux qui s'intéressent le plus aux nouvelles technologies, ou ceux qui sont le plus motivés;
- faire appel à différents utilisateurs à chaque nouvelle session de tests pour éviter les effets d'apprentissage.

L'interface

Les tests d'utilisabilité sont possibles à partir du moment où l'on dispose d'un premier prototype même incomplet de l'interface. Avant ce stade, on peut solliciter l'avis de sujets sur différentes versions papier de l'interface, mais il s'agit d'un exercice plutôt abstrait, qui est

encore loin de la réalité. Les tests peuvent se poursuivre avec des prototypes de plus en plus complets et formels jusqu'à ce que l'on soit satisfait de l'ensemble de l'interface, incluant le manuel de l'opérateur et la documentation qui doivent aussi être testés.

Il faut décrire, même sommairement, les caractéristiques de l'interface et du système informatique sur lequel elle est installée. D'abord, parce que l'interface peut changer entre le moment où l'on décide de faire des tests et celui où l'on présente les résultats, et on a besoin d'un point de référence stable pour situer et décrire les problèmes de l'interface. Puis, parce que le système sur lequel l'interface est installée peut être différent de celui ayant servi au développement, et la performance de l'interface peut être affectée. La description de l'interface doit être faite du point de vue de l'utilisateur, non du point de vue informatique, afin de montrer ce que l'on voit, entend et manipule. Concernant l'interface, on va noter le numéro de la version, l'outil de prototypage utilisé et, pour éviter des descriptions laborieuses, on va montrer les copies des principales pages-écrans et de celles mentionnées dans les résultats. Concernant le système informatique, on va décrire la plate-forme informatique (incluant le microprocesseur, le clavier, la souris, l'écran, le microphone, la caméra et les haut-parleurs, s'il y a lieu), le système d'exploitation, l'architecture globale du système et les bases de données auxquelles on accède.

Pour mener à bien son travail, le responsable des tests doit avoir une bonne connaissance des caractéristiques de l'interface à tester, afin de pouvoir choisir des tâches qui sont possibles compte tenu du niveau de développement de l'interface, comprendre les stratégies et les difficultés des utilisateurs et pouvoir les analyser à la lumière de ce que propose l'interface.

La procédure des tests

Il est important de définir le protocole des tests de façon complète et détaillée afin de ne rien oublier, de suivre la même démarche avec chacun des utilisateurs participants et de recueillir les données de façon systématique. Ainsi, les points suivants doivent être bien décrits:

- l'organisation de la salle de tests comprenant le système informatique, l'équipement audio-visuel (ex., caméras, microphones), les manuels d'aide, etc.;
- la liste des tâches à effectuer;
- les consignes de travail, comprenant le but et le déroulement des tests, les tâches à effectuer, ce que l'on attend de l'utilisateur et la procédure à suivre (ex., "penser tout haut", possibilité de demander de l'aide après des tentatives infructueuses);
- les données à recueillir et la méthode de cueillette pour chacune: par ex., faire une entrevue, prendre des notes (ex., au sujet des comportements de l'utilisateur), enregistrer les commentaires verbaux, remplir un questionnaire, faire l'enregistrement vidéo de la session.

Si possible, une session de tests ne devrait pas excéder une heure afin d'éviter les problèmes de fatigue chez les utilisateurs; ne pas dépasser un maximum de deux heures, incluant une pause d'environ 10 minutes.

Les mesures. Le tableau 1 dresse la liste des paramètres de la performance et de la satisfaction humaine qui peuvent être mesurés au cours des tests d'utilisabilité (à noter qu'ils sont en grande majorité objectifs et quantifiables). Cette liste recouvre celle proposée par d'autres auteurs (ex., Nielsen 1994a). Le problème du choix des paramètres pertinents ressemble à plusieurs égards à celui qui se pose dans la recherche sur les facteurs humains (Kantovitz, 1992). Il dépend de plusieurs facteurs, dont les objectifs généraux de la session de tests, les objectifs de chaque tâche ou scénario, la phase de conception, la criticalité des tâches, les contraintes de temps et de budget, et parfois les restrictions imposées par les utilisateurs, le syndicat ou l'entreprise.

En analysant ces différents facteurs et de concert avec l'équipe de conception, le responsable des tests parvient à définir des objectifs à atteindre pour chaque tâche ou scénario de tâche: par ex., 80% des utilisateurs doivent réussir cette tâche en 1 ou 2 tentatives, 100% des utilisateurs

Tableau 1: Paramètres de la performance et de la satisfaction humaine pouvant être mesurés au cours des tests d'utilisabilité (tirés de Preece et al., 1994)

- le temps d'exécution de la tâche
- le pourcentage complété de la tâche
- le pourcentage complété de la tâche par unité de temps
- le taux de succès
- le nombre d'erreurs
- le temps consacré aux erreurs
- la gravité des erreurs
- le pourcentage ou le nombre de systèmes compétitifs qui font mieux cette tâche que le produit actuel
- le nombre de commandes utilisées vs inutilisées
- la fréquence de l'utilisation de l'aide ou de la documentation
- le temps passé à utiliser l'aide ou à lire la documentation
- le nombre d'utilisations répétées de commandes ayant échoué
- le nombre de fois que que l'interface a mal guidé l'utilisateur
- le nombre de retours en arrière
- le nombre de fois que l'utilisateur a dû revenir en arrière
- le nombre de fois que l'utilisateur a digressé de sa tâche
- le nombre de fois que l'utilisateur a perdu le contrôle du système
- le nombre de fois que l'utilisateur a exprimé de la satisfaction ou de la frustration
- le nombre de bons et mauvais aspects du système dont les utilisateurs se rappellent
- le nombre de commentaires favorables vs défavorables des utilisateurs
- le niveau de satisfaction à l'égard du système

doivent réussir cette tâche (critique) sans erreur et sans aide. Ces objectifs indiquent les paramètres pertinents à choisir et les niveaux à atteindre pour chacun d'eux.

D'autres mesures sont possibles plus en amont de celles liées à la performance et à la satisfaction des utilisateurs. Elles ont trait aux problèmes de l'interface qui sont réputés influer sur ces deux dernières. Ces problèmes mettent en cause l'un ou l'autre des facteurs d'utilisabilité des interfaces qui ont été définis par différents auteurs (ex., Bastien & Scapin, 1994; Ravden & Johnson, 1989; Robert, 1990) (voir tableau 2). Voici des exemples de problèmes de l'interface:

- les noms des commandes et des menus manquent de cohérence;
- la navigation de tel écran à tel écran est laborieuse, elle fait perdre du temps;
- la signification d'une icone n'est pas claire;
- des messages d'erreurs sont incompréhensibles;
- il n'y a pas de retour d'information pour une certaine action;
- il n'y a pas d'aide en ligne pour un certain problème;
- les touches *Help* et *Delete* sont trop rapprochées l'une de l'autre, cela occasionne des erreurs.

RESULTATS DE NOS PROJETS

Le laboratoire d'utilisabilité ICONE a permis de tester des interfaces graphiques surtout destinées à des clientèles internes des entreprises, une seule interface ayant été destinée au grand public. Ces interfaces devaient être utilisées dans les environnements de bureaux, dans les télécommunications ou en médecine. Leur degré de complexité va de simple (ex., quelques

Tableau 2: Facteurs d'utilisabilité des interfaces humain-machine

Bastien & Scapin (1993)	**Ravden & Johnson (1989)**	**Robert (1990)**
1. Guidage 2. Charge de travail 3. Contrôle explicite 4. Adaptabilité 5. Gestion des erreurs 6. Homogénéité/Cohérence 7. Signifiance des codes et dénomination 8. Compatibilité	1. Clarté visuelle 2. Cohérence 3. Compatibilité 4. Bon retour d'information 5. Caractère explicite 6. Fonctionnalités appropriées 7. Flexibilité et contrôle 8. Prévention et correction des erreurs 9. Guidage et soutien à l'utilisateur	Fonctionnalités appropriées **Qualités de l'interface:** - Compatibilité - Caractère naturel - Transparence - Cohérence - Contrôlabilité - Rapidité (d'actions) - Simplicité - Flexibilité/Adaptabilité **Fonctionnalités*:** - Retour d'information - Navigation - Aide - Gestion des erreurs

* Plusieurs des qualités mentionnées plus haut s'appliquent à ces fonctionnalités.

pages-écrans) à très complexe. Les paragraphes qui suivent présentent plusieurs constatations importantes sur le rôle et le mode de fonctionnement du laboratoire.

Avoir un laboratoire portatif pouvant facilement être installé sur le terrain.
Tous les tests d'utilisabilité ont été réalisés sur le terrain, c.-à-d. dans le milieu de travail des futurs utilisateurs ou parfois chez le développeur de l'application, mais jamais en laboratoire. Le laboratoire doit donc être portatif afin de pouvoir être déplacé et installé facilement dans diverses entreprises. Les avantages de procéder ainsi sont nombreux:

- les utilisateurs sont dans leur milieu, ce qui est plus naturel et moins stressant pour eux;
- les utilisateurs sont plus facilement disponibles pour participer aux tests parce qu'ils sont sur place et qu'ils savent qu'ils seront moins longtemps absents de leur travail; en même temps, les frais encourus par leur participation sont plus limités;
- on peut faire les tests avec le système informatique de l'entreprise.

Le principal inconvénient porte sur la difficulté de contrôler parfaitement les conditions de tests, car malgré les consignes, des utilisateurs continuent parfois de prendre leurs appels téléphoniques ou de répondre à des collègues, et nous n'avons pas le plein contrôle pas sur les renseignements qu'ils pourraient se communiquer au sujet des tests.

Prévoir plusieurs sessions de tests au cours de la conception.
Les tests d'utilisabilité s'inscrivent dans une démarche itérative de test, correction et retest de l'interface. Il faut donc s'attendre à devoir faire plusieurs sessions de tests au cours de la conception, ce que nos différents projets ont confirmé. Le nombre de sessions nécessaires dépend de l'écart qui existe entre la qualité de la première version de l'interface qui a été réalisée et celle que l'on vise comme produit final, de l'ampleur du travail de reconception qui a été fait après chaque session de tests, en plus des contraintes de temps et de budget. A titre de référence, mentionnons que Gould et al. (1987) prétendent que les guides de l'utilisateur du système de messagerie (fait par IBM) pour les Jeux olympiques de 1984 ont fait l'objet de 200

itérations avant d'être distribués aux Jeux olympiques!

Avant les tests d'utilisabilité, faire une évaluation heuristique de l'interface.
Avec l'engouement actuel pour les tests d'utilisabilité et leur apparente simplicité (d'aucuns croient qu'il suffit de demander aux utilisateurs), en plus de la méconnaissance des différentes méthodes d'évaluation des interfaces dans les milieux industriels, on nous demande souvent de commencer l'évaluation par ces tests. Nous estimons que c'est une perte de temps et d'argent lorsque les failles de l'interface sautent aux yeux. Une évaluation heuristique plus rapide et moins coûteuse serait plus efficace. Malgré cela, il demeure très important d'associer les utilisateurs le plus tôt possible au processus de conception et d'évaluation de l'interface.

Au cours des premiers tests, rechercher les grands problèmes de l'interface.
Tout test d'utilisabilité porte au moins sur un aspect de la performance de l'utilisateur, ne serait-ce que parce qu'on vérifie d'abord si la tâche a pu être réalisée, et parfois aussi en combien de temps. Mais en début de conception, l'étude de la performance n'est pas une fin en soi, c'est plutôt un moyen de découvrir les failles de l'interface. Les tests d'utilisabilité que nous avons réalisés se sont toujours concentrés en premier lieu sur les grands problèmes de l'interface, concernant surtout sa compatibilité avec l'organisation de la tâche et avec les besoins et les attentes des utilisateurs. On cherchait d'abord à éliminer les problèmes avant de mesurer formellement la performance et la satisfaction des utilisateurs.

Se servir des tests d'utilisabilité pour compléter l'analyse des besoins.
Les tests d'utilisabilité se révèlent être un excellent moyen d'analyser les besoins et les attentes des utilisateurs et de compléter ainsi l'analyse traditionnelle des besoins préalable à un système. Grâce à des logiciels de prototypage très performants, on peut construire rapidement et très tôt dans le processus de conception des prototypes d'interfaces avec lesquels les utilisateurs peuvent interagir. Ainsi, parce qu'ils peuvent faire des tâches concrètes dans un mode interactif et voir les possibilités offertes par la technologie, les utilisateurs peuvent exprimer leurs besoins et leurs attentes de façon beaucoup plus précise et plus concrète qu'auparavant.

Se servir des tests d'utilisabilité pour analyser l'activité des utilisateurs.
Les tests d'utilisabilité se révèlent aussi être un excellent moyen d'analyser l'activité des utilisateurs au cours de leurs interactions avec le système. Cette analyse est différente de celle qui est faite préalablement à la réalisation de l'interface. Elle prend le relais et compense les lacunes de cette analyse préalable car plus l'analyse préalable est incomplète, plus on court le risque de construire une première interface médiocre qui devra être testée, modifiée et retestée plusieurs fois, occasionnant ainsi plusieurs cycles d'analyse de l'activité de l'utilisateur avec l'interface. Cette analyse est nécessaire pour l'équipe de conception parce qu'elle montre comment l'interface est réellement utilisée et quelles difficultés elle engendre.

Recueillir toute l'information pertinente au moment des tests.
Pendant les tests, on a demandé à l'utilisateur qui travaille seul de "penser tout haut" afin de mieux connaître ses stratégies de travail, ses difficultés et ses impressions; les commentaires verbaux étaient enregistrés. Jusqu'à maintenant, on a fait des observations directes dans la salle de tests en notant les comportements et les attitudes de l'utilisateur. Mais cette façon de faire crée de l'embarras chez certains sujets de sorte qu'il vaudrait mieux observer indirectement à travers une console vidéo ou un terminal esclave relié au système en opération. On insiste aussi pour qu'un membre de l'équipe de conception assiste aux tests d'utilisabilité afin d'être plus sensibilisé aux problèmes de l'interface. On fait une entrevue synthèse avec chaque sujet à la fin des tests. On procède à l'enregistrement vidéo de la session dans la majorité des cas afin de garder une trace des résultats et de pouvoir montrer des extraits aux concepteurs, s'il y a lieu. On fait toutefois peu usage de ces enregistrements pour l'analyse en raison des coûts. De plus, l'enregistrement vidéo s'avère peu utile lorsqu'on l'on traite les données immédiatement, sans compter qu'il alourdit les tests sur le terrain à cause de l'équipement à installer et à surveiller.

Par ailleurs, les utilisateurs et les syndicats sont parfois réticents face à ces enregistrements car ils y voient une forme d'évaluation de la performance.

Les tests ne permettent pas de détecter tous les problèmes de l'interface.
Le faible nombre de sujets qui participent aux tests ne permet pas de détecter tous les problèmes de l'interface. Même lorsque ce nombre est plus élevé, les tests d'utilisabilité s'avèrent inefficaces pour trouver les problèmes bien cachés ou n'ayant pas de répercussions immédiates sur la performance et la satisfaction de l'utilisateur. Des problèmes tels que le manque de conformité à des normes, des incohérences dans la terminologie et les nombreux messages d'aide, des procédures non optimisées, des fautes de frappe ou d'orthographe dans les messages, un mauvais choix de couleurs, etc. Nous avons donc dû compléter les tests par d'autres formes d'évaluation plus systématiques.

Les tests mesurent la facilité d'utilisation à court terme et ne font pas de suivi.
Parce que chaque session de tests est très brève (ex., maximum de 2 heures par sujet) et que l'on change de sujets d'une session à l'autre, les tests correspondent à une étude transversale et non pas longitudinale de l'interface. L'on sait peu de choses sur l'apprentissage et l'utilisation de l'interface à moyen et à long termes (ex., après 6 mois d'usage intense). En réalité, on s'intéresse aux premières réactions d'un utilisateur novice travaillant en mode d'exploration et on juge l'interface de façon très conservatrice en statuant qu'elle est facile à utiliser si le sujet réussit du premier coup. Cela n'est qu'une facette de l'interaction avec un système.Il faudrait faire une étude de suivi sur la performance et la satisfaction à moyen et à long termes.

Proposer une solution pour chaque problème et le plus tôt possible après les tests.
Les tests font toujours l'objet d'un rapport écrit qui est parfois accompagné d'une présentation verbale. Il s'agit d'un rapport technique classique qui décrit le contexte et les objectifs de l'étude, la méthodologie suivie -incluant les tâches, les utilisateurs, l'interface, la procédure de tests, les outils, les contraintes de temps et de budget et les limites de l'étude-, les résultats, des recommandations, une conclusion et des références bibliographiques. Le rapport doit décrire les problèmes de l'interface et leurs interrelations, montrer leurs impacts sur les utilisateurs, évaluer leur degré d'importance et donner un ordre de priorité pour la correction, et proposer une solution pour chaque problème. Il doit aussi présenter les résultats des mesures de performance et de satisfaction des utilisateurs pour chaque tâche, dire si les objectifs ont été atteints et si non, expliquer pourquoi. Les résultats personnels doivent être présentés de façon anonyme. Le rapport doit être remis le plus tôt possible après les tests, dans un délai de quelques jours.

CONCLUSION

Au terme d'une première année d'opération du laboratoire ICONE, on constate que les tests d'utilisabilité ont tous été effectués sur le terrain et nécessitent donc des équipements portatifs. Plusieurs sessions de tests ont été nécessaires pour chaque projet; au tout début, elles portaient sur les problèmes de l'interface, et par la suite elles visaient à mesurer la performance et la satisfaction des utilisateurs. Les tests ont surtout permis de mesurer la facilité d'utilisation de l'interface à court terme puisqu'on a fait appel à des sujets novices travaillant en mode d'exploration, et qu'on a jugé l'interface essentiellement à partir des premières réactions des utilisateurs. Enfin, les tests se sont avérés une excellente façon de compléter l'analyse des besoins et d'analyser l'activité des utilisateurs dans leurs interactions avec le système.

Trois pistes de recherche nous paraissent prometteuses pour améliorer les tests d'utilisabilité: mesurer la représentativité des différents scénarios de tâches soumis aux utilisateurs, développer une méthode formelle pour évaluer rapidement et sûrement la gravité des problèmes rencontrés dans l'interface, et investiguer l'activité de conception ou de reconception de l'interface qui fait suite aux problèmes soulevés par chaque session de tests.

REFERENCES

Bastien, J.M.C., Scapin, D.L. (1993). Critères ergonomiques pour l'évaluation d'interfaces. Utilisateurs. Rapport technique # 156, Programme 3 Intelligence artificielle, Systèmes cognitifs et Interaction homme-machine, INRIA, France.

Centre de compétence en ergonomie. Banc d'essai - Interfaces Utilisateurs, Dépliant de promotion, Bayonne, France, 1993.

Chignell, M. H. (1992). Usability laboratories (Université de Toronto). Communiqué, 23 (1), 1-2.

Gould, J.D. (1988). How to design usable systems. In Helander, M. (Ed.). Handbook of human-computer interaction. Amsterdam, Elsevier, North-Holland, p. 757-789.

Gould, J.D., Boies, S.J., Levy, S., Richards, J.T., Shoonard, J. (1987). The 1984 olympic message system: A test of behavioral principles of system design. Communications of the ACM, 30, 758-769.

ICONE (1994). Dépliant d'information sur le laboratoire d'utilisabilité ICONE. Ecole Polytechnique de Montréal et Centre de Recherche en Informatique de Montréal, Québec.

Kantovitz, B.H. (1992). Selecting measures for human factors research. Human Factors, 34, 387-398.

Karat, J. (1988). Software evaluation methodologies. In Helander, M. (Ed.). Handbook of human-computer interaction. Amsterdam, Elsevier, North-Holland, p. 891-903.

LablUtil (1995). Laboratoire d'utilisabilité de l'Université fédérale de Santa Catarina à Florianopolis et du Centre de Technologie en Automation et en Informatique, Brésil.

Mack, R.L., Nielsen, J. (1993). Usability inspection methods. ACM SIGCHI Bulletin, 25, 1, 28-33.

Nielsen, J. (Ed.) (1994a). Usability laboratories. Behaviour & Information Technology, 13, 2, March-April.

Nielsen, J. (1994b). Usability engineering. AP Professional. Harcourt Brace & Company, Boston.

Mayhew, D.J. (1992). Principles and guidelines in software user interface design. Prentice Hall.

Preece, J., Rogers, Y., Sharp, H., Benyon, D.H., Golland, S., Carey, T. (1994). Human-computer interaction. Addison-Wesley, Reading, MA.

Ravden, S.J., Johnson, G.I. (1989). Evaluating usability of human-computer interfaces. A practical method. Chichester: Ellis Horwood.

Reed, S. (1992). Who defines usability? You do. PC Computing, December, 220-232.

Robert, J.-M. (1990). Les facteurs d'utilisabilité des interfaces humain-machine. Notes de cours, Ecole Polytechnique de Montréal. Document inédit.

Robert, J.-M., Fiset, J.-Y. (1992). Conception et évaluation ergonomiques d'une interface pour un logiciel d'aide au diagnostic; une étude de cas. ICO, no 2, p. 67-73.

Scerbo, M.W. (1995). Usability testing. In Weimer, J. (Ed.). Research techniques in human engineering. Prentice Hall, Englewood Cliffs, N.J., p. 72-111,

Tullis, T.S. (1993). Is user interface design just common sense? In Proceedings of the HCI International'93 - 5th Conference on Human-Computer Interaction, 8 - 13 août, Orlando, Floride, p. 9-14.

Virzi, R.A. (1992). Refining the test phase of usability evaluation: How many subjects is enough? Human Factors, 34, 457-468.

Whiteside, J., Bennett, J., Holtzblatt, K. (1988). Usability engineering: Our experience and evolution. In Helander, M. (Ed.). Handbook of human-computer interaction. Amsterdam: Elsevier, North-Holland, p. 791-817.